1. Record Nr. UNINA9910485588003321

Autore	Shiyal Bhadresh
Titolo	Beginning Azure synapse analytics : transition from data warehouse to data lakehouse / / Bhadresh Shiyal
Pubbl/distr/stampa	[Place of publication not identified] : , : Apress, , [2021] ©2021
ISBN	1-4842-7061-4
Descrizione fisica	1 online resource (263 pages)
Disciplina	658.40380285574
Soggetti	Data warehousing - Management Microsoft Azure (Computing platform)
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Table of Contents -- About the Author -- About the Technical Reviewer -- Acknowledgments -- Introduction -- Chapter 1: Core Data and Analytics Concepts -- Core Data Concepts -- What Is Data? -- Structured Data -- Semi-structured Data -- Unstructured Data -- Data Processing Methods -- Batch Data Processing -- Streaming or Real-Time Data Processing -- Relational Data and Its Characteristics -- Non-Relational Data and Its Characteristics -- Core Data Analytics Concepts -- What Is Data Analytics? -- Data Ingestion -- Data Exploration -- Data Processing -- ETL -- ELT -- ELT / ETL Tools -- Data Visualization -- Data Analytics Categories -- Descriptive Analytics -- Diagnostic Analytics -- Predictive Analytics -- Prescriptive Analytics -- Cognitive Analytics -- Summary -- Chapter 2: Modern Data Warehouses and Data Lakehouses -- What Is a Data Warehouse? -- Core Data Warehouse Concepts -- Data Model -- Model Types -- Schema Types -- Metadata -- Why Do We Need a Data Warehouse? -- Efficient Decision-Making -- Separation of Concerns -- Single Version of the Truth -- Data Restructuring -- Self-Service BI -- Historical Data -- Security -- Data Quality -- Data Mining -- More Revenues -- What Is a Modern Data Warehouse? -- Difference Between Traditional &amp -- Modern Data Warehouses -- Cloud vs. On-Premises -- Separation of Compute and Storage Resources -- Cost -- Scalability -- ETL vs. ELT -- Disaster Recovery -- Overall Architecture -- Data Lakehouse --

What Is a Data Lake? -- What Is Delta Lake? -- What Is Apache Spark? -- What Is a Data Lakehouse? -- Characteristics of a Data Lakehouse -- Various Data Types -- AI -- Decoupled Compute and Storage Resources -- Open Source Storage Format -- Data Analytics and BI Tools -- ACID Properties -- Differences Between a Data Warehouse and a Data Lakehouse -- Architecture -- Access to Raw Data.
Open Source vs. Proprietary -- Workloads -- Query Engines -- Data Processing -- Real-Time Data -- Examples of Data Lakehouses -- Azure Synapse Analytics -- Databricks -- Benefits of Data Lakehouse -- Support for All Types of Data -- Time to Market -- More Cost Effective -- AI -- Reduction in ETL/ELT Jobs -- Usage of Open Source Tools and Technologies -- Efficient and Easy Data Governance -- Drawbacks of Data Lakehouse -- Monolithic Architecture -- Technical Infancy -- Migration Cost -- Lack of Many Products/Options -- Scarcity of Skilled Technical Resources -- Summary -- Chapter 3: Introduction to Azure Synapse Analytics -- What Is Azure Synapse Analytics? -- Azure Synapse Analytics vs. Azure SQL Data Warehouse -- Why Should You Learn Azure Synapse Analytics? -- Main Features of Azure Synapse Analytics -- Unified Data Analytics Experience -- Powerful Data Insights -- Unlimited Scale -- Security, Privacy, and Compliance -- HTAP -- Key Service Capabilities of Azure Synapse Analytics -- Data Lake Exploration -- Multiple Language Support -- Deeply Integrated Apache Spark -- Serverless Synapse SQL Pool -- Hybrid Data Integration -- Power BI Integration -- AI Integration -- Enterprise Data Warehousing -- Seamless Streaming Analytics -- Workload Management -- Advanced Security -- Summary -- Chapter 4: Architecture and Its Main Components -- High-Level Architecture -- Main Components of Architecture -- Synapse SQL -- Compute Layer -- Dedicated Synapse SQL Pool -- Serverless Synapse SQL Pool -- Storage Layer -- Synapse Spark or Apache Spark -- Synapse Pipelines -- Synapse Studio -- Synapse Link -- Summary -- Chapter 5: Synapse SQL -- Synapse SQL Architecture Components -- Massively Parallel Processing Engine -- Distributed Query Processing Engine -- Control Node -- Compute Nodes -- Data Movement Service -- Distribution -- Hash Distribution.
Round-Robin Distribution -- Replication-based Distribution -- Azure Storage -- Dedicated or Provisioned Synapse SQL Pool -- Serverless or On-Demand Synapse SQL Pool -- Synapse SQL Feature Comparison -- Database Object Types -- Query Language -- Security -- Tools -- Storage Options -- Data Formats -- Resource Consumption Model for Synapse SQL -- Synapse SQL Best Practices -- Best Practices for Serverless Synapse SQL Pool -- Best Practices for Dedicated Synapse SQL Pool -- How-To's -- Create a Dedicated Synapse SQL Pool -- Create a Serverless or On-Demand Synapse SQL Pool -- Load Data Using COPY Statement in Dedicated Synapse SQL Pool -- Ingest Data into Azure Data Lake Storage Gen2 -- Summary -- Chapter 6: Synapse Spark -- What Is Apache Spark? -- What Is Synapse Spark in Azure Synapse Analytics? -- Synapse Spark Features &amp -- Capabilities -- Speed -- Faster Start Time -- Ease of Creation -- Ease of Use -- Security -- Automatic Scalability -- Separation of Concerns -- Multiple Language Support -- Integration with IDEs -- Pre-loaded Libraries -- REST APIs -- Delta Lake and Its Importance in Synapse Spark -- Synapse Spark Job Optimization -- Data Format -- Memory Management -- Data Serialization -- Data Caching -- Data Abstraction -- Join and Shuffle Optimization -- Bucketing -- Hyperspace Indexing -- Synapse Spark Machine Learning -- Data Preparation and Exploration -- Build Machine Learning Models -- Train Machine Learning Models -- Model Deployment and Scoring -- How-To's --

How to Create a Synapse Spark Pool -- How to Create and Submit Apache Spark Job Definition in Synapse Studio Using Python -- How to Monitor Synapse Spark Pools Using Synapse Studio -- Summary -- Chapter 7: Synapse Pipelines -- Overview of Azure Data Factory -- Overview of Synapse Pipelines -- Activities -- Pipelines -- Linked Services -- Dataset -- Integration Runtimes (IR).
Azure Integration Runtime (Azure IR) -- Self-Hosted Integration Runtimes (SHIR) -- Azure SSIS Integration Runtimes (Azure SSIS IR) -- Control Flow -- Parameters -- Data Flow -- Data Movement Activities -- Category: Azure -- Category: Database -- Category: NoSQL -- Category: File -- Category: Generic -- Category: Services and Applications -- Data Transformation Activities -- Control Flow Activities -- Copy Pipeline Example -- Transformation Pipeline Example -- Pipeline Triggers -- Summary -- Chapter 8: Synapse Workspace and Studio -- What Is a Synapse Analytics Workspace? -- Synapse Analytics Workspace Components and Features -- Azure Data Lake Storage Gen2 Account and File System -- Serverless Synapse SQL Pool -- Shared Metadata Management -- Code Artifacts -- What Is Synapse Studio? -- Main Features of Synapse Studio -- Home Hub -- Data Hub -- Develop Hub -- Integrate Hub -- Monitor Hub -- Integration -- Activities -- Manage Hub -- Analytics Pools -- External Connections -- Integration -- Security -- Synapse Studio Capabilities -- Data Preparation -- Data Management -- Data Exploration -- Data Warehousing -- Data Visualization -- Machine Learning -- Power BI in Synapse Studio -- How-To's -- How to Create or Provision a New Azure Synapse Analytics Workspace Using Azure Portal -- How to Launch Azure Synapse Studio -- How to Link Power BI with Azure Synapse Studio -- Summary -- Chapter 9: Synapse Link -- OLTP vs. OLAP -- What Is HTAP? -- Benefits of HTAP -- No-ETL Analytics -- Instant Insights -- Reduced Data Duplication -- Simplified Technical Architecture -- What Is Azure Synapse Link? -- Azure Cosmos DB -- Azure Cosmos DB Analytical Store -- Columnar Storage -- Decoupling of Operational Store -- Automatic Data Synchronization -- SQL API and MongoDB API -- Analytical TTL -- Automatic Schema Updates -- Cost-Effective Archiving -- Scalability.
When to Use Azure Synapse Link for Cosmos DB -- Azure Synapse Link Limitations -- Azure Synapse Link Use Cases -- Industrial IOT -- Predictive Maintenance Pipeline -- Operational Reporting -- Real-Time Applications -- Real-Time Personalization for E-Commerce Users -- How-To's -- How to Enable Azure Synapse Link for Azure Cosmos DB -- How to Create an Azure Cosmos DB Container with Analytical Store Using Azure Portal -- How to Connect to Azure Synapse Link for Azure Cosmos DB Using Azure Portal -- Summary -- Chapter 10: Azure Synapse Analytics Use Cases and Reference Architecture -- Where Should You Use Azure Synapse Analytics? -- Large Volume of Data -- Disparate Sources of Data -- Data Transformation -- Batch or Streaming Data -- Where Should You Not Use Azure Synapse Analytics? -- Use Cases for Azure Synapse Analytics -- Financial Services -- Manufacturing -- Retail -- Healthcare -- Reference Architectures for Azure Synapse Analytics -- Modern Data Warehouse Architecture -- Real-Time Analytics on Big Data Architecture -- Summary -- Index.